JUDAISM, SCIENCE, AND MORAL RESPONSIBILITY

Edited by
Yitzhak Berger and
David Shatz

ROWMAN & LITTLEFIELD PUBLISHERS, INC.
Lanham • Boulder • New York • Toronto • Oxford
1

A Scientific Perspective on Human Choice

Haim Sompolinsky

I. The Problem

We are used to thinking about ourselves as having the freedom to make decisions that affect the course of our lives to a significant degree. We regard our ability to make such choices as a key characteristic of the human condition. At the same time, we are part of a universe that appears to operate in fairly fixed and unalterable ways. This gives rise to what philosophers call the problem of free choice (or free-will) and determinism: what does it mean for us humans to be truly free in a world where the future is determined by the present? Do we possess free will, or are all our actions and mental states inexorably determined?

The notion of free will is one of the most controversial issues in philosophy, and determinism is just one of the difficulties surrounding it. Equally problematic is the role of our will or choice in an indeterministic world, since randomness and chance seem to be blind to the power of our will. In this article, I will sketch the broad lines of a scientific perspective of human choice. In Section II, I discuss the notions of determinism and predictability in classical and modern physics. Section III outlines current research on the neuronal processes that underlie human cognition and choice. In Section IV, I discuss the main attempts to reconcile the strong sense of free will—metaphysical freedom of the will—with scientific conceptions and knowledge. Section V is devoted to a critical appraisal of these attempts, emphasizing the problem of will/brain interaction. Summary and concluding remarks are presented in Section VI.
II. Determinism and Probabilities in the Laws of Physics

1. Determinism in Classical Physics

The problem of free will is commonly referred to as the conflict between the concept of free will and the concept of determinism. There are various kinds of determinism, including logical, physical, and theological determinism. In this paper, I focus on physical determinism.

A fundamental consequence of the principles of classical physics is that physical laws fully determine the way events in nature unfold. According to classical theory, all future states of a physical system are determined by its present state and the forces acting on it. Let us consider, for example, an isolated mechanical system. The physical state of the system at any time consists of the values of the positions and velocities of all the particles comprising the system at that time. These particles are acted upon by gravitational, mechanical, or electrostatic forces. According to Newton’s Laws of Mechanics, the positions and instantaneous velocities of all the particles at a given time, together with the forces between them, fully determine the future (and the past) movements of all the particles in the system.

The above principle of determinism applies to an isolated system. In reality, physical systems are in continuous contact with their environment and exchange momentum and energy with it. In such a situation, the principle of determinism still holds, but the interaction of the system with its environment should be included among the forces that affect the motion of the particles in the system.

An alternative exposition of the principle of determinism, which avoids this complication, is to apply it to the entire universe. This giant conceptual step was made by the French mathematician and philosopher Pierre Simon de Laplace. In his Philosophical Essay on Probability (1819), Laplace argues that all natural events are connected to previous ones by universal causation: “All events, even those which, because of their small scale, do not appear to keep to the great laws of nature, are just as necessary a result of those laws as are the revolutions of the sun.” In other words, there should be a set of scientific laws that characterize the physical forces at all scales of matter, and these laws together with the state of the universe at a given time will fully determine the future course of the entire universe.

What do we mean by claiming that the laws of nature “fully determine” the future course of events? Essentially, we mean that the laws of nature are sufficient “causes” of future events; namely that the future state of a physical system is necessitated by its present state and the physical forces operating on it.
A Scientific Perspective on Human Choice

The concept of causality has been the subject of controversy for centuries and an exposition of the myriad of attendant metaphysical questions is beyond the scope of this paper. However, in the present context there is a simple operational definition of “full determination” of the state of a system. If we have full knowledge of the initial state of a physical system and the forces that operate on it we can calculate its future state. In other words, classical determinism here implies predictability, at least in theory. This relationship between determinism and predictability was emphasized by Laplace himself:

We must therefore regard the present state of the universe as the effect of its preceding state and as the cause of the one which is to follow. An intelligence which in a singular instant could know all the forces which animate the natural world, and the respective situations of all the beings that made it up, could, provided it was vast enough to make an analysis of all the data so supplied, be able to produce a single formula which specified all the movements in the universe from those of the largest bodies in the universe to those of the lightest atom. For such intelligence, nothing would be “uncertain,” and the future, like the past, would be present before its eyes.

Laplacian determinism projects a naïve view of physical systems as clock-like predictable machines. Modern physics provides a much more complex picture of determinism and predictability. In the following section I will briefly review some of the new perspectives provided by modern physics.

2. Chaos and the Limits of Predictability of Physical Systems

Laplace’s equating of determinism with predictability raises the question of the feasibility of determining the initial state of a system with the accuracy required to predict its future state. Of course, Laplace’s super-intelligent demon that has full knowledge of the initial condition of the entire universe is an outright fiction. The more rigorous issue is this: how precisely must we know the initial state of a physical system in order to accurately predict its future course?

Obviously, we can never know the state of any system with infinite accuracy. Nevertheless, our daily experiences inform us that we can often make reliable predictions about the future course of natural events. Apparently, some systems’ future behavior is not very sensitive to mild changes in their initial condition; hence, we do not need to have very precise information about their initial state in order to predict their future state. Are all physical systems like this?
The modern theory of chaos (originally developed by Henri Poincaré around the turn of the twentieth century) has revealed that many physical systems (e.g., air currents in the atmosphere, a dripping faucet, electronic circuits, heart beats) are so exquisitely sensitive that the slightest disturbance in their initial conditions can affect their behavior radically. The degree of unpredictability depends on the degree of chaos in the system and the length of the attempted prediction. Even in a chaotic system, short-term behavior can often be predicted reliably. Hence, chaos does not imply a complete loss of predictability. Rather, it implies that the error in predicting the future course grows extremely fast with the time elapsed since the state of the system was determined to a given degree of accuracy. The further we want to predict into the future, the more accurately we have to determine the initial state. The amount of accuracy required grows so fast with the time of prediction that long-term prediction is practically impossible for chaotic systems. The weather is an example of a chaotic system; because of its chaotic nature, long-term precise forecasting is virtually impossible. Nevertheless, modern observation and modeling techniques allow for relatively reliable short-term weather forecasting.

This analysis makes it clear that in chaotic systems, it is predictability that really breaks down, not the cause-effect chain of events. A chaotic system is no less determined by the physical laws than a nonchaotic system—it is simply more difficult to predict its course over long periods. This is why the more accurately we know the initial state of a chaotic system, the better our prediction of its behavior will be. The fact that there is no theoretical limit on the accuracy of determination of the initial state of a chaotic system (as opposed to practical limits) implies that there is no theoretical limit on the ability to predict its future.

3. Statistical Physics

The range of phenomena in the real world that are described efficiently by deterministic laws such as the Newtonian laws of mechanics is very limited. Very often the properties of macroscopic systems at ordinary temperatures are not described by these laws but by statistical laws, formulated by a branch of physics called statistical mechanics. At the core of statistical mechanics lies the distinction between the microstate of a large system and its macrostate. The microstate is specified by the positions and velocities of each atom and molecule, as in classical mechanics. The macrostate corresponds to a few macroscopic properties, such as the system’s volume, pressure, and temperature. Rather than attempting to predict the detailed microstate of the system, statistical mechanics establishes the probability that the system will be in each
of its different microstates, and the relation between these probabilities and the macrostate of the system.

It is an important and interesting fact that most of our accounts of macroscopic systems, including biological ones, are far removed from the deterministic microscopic laws of motion, the kind of laws that Laplace had in mind. Basic physical concepts, such as heat and temperature, are grounded in the statistical description of the states of the system. Noise, randomness, fluctuations, and probabilities dominate the terrain of modern physics. One may wonder about the source of this probabilistic description of nature. What is the justification for ignoring the deterministic picture of microscopic classical physics when dealing with macroscopic objects?

Although a full account of how the probabilistic laws of statistical mechanics arise from the deterministic laws of microphysics may still be incomplete, the general picture is clear. Ordinary macroscopic systems are not isolated. They are in continuous interaction with the environment. The microstate of the system is sensitive to the details of the extremely complex microscopic interactions between the system and its environment. Often, the influence of the environment on the surrounded system takes the form of a force whose value fluctuates strongly in space and time. These fluctuations are called thermal fluctuations. The higher the temperature of the system (and its environment), the larger these fluctuations become. However, the macroscopic properties of the system are relatively insensitive to the details of its microstate, and are therefore insensitive to the details of the system-environment interactions. Furthermore, because of the interaction with the environment, the microstate of the system changes very rapidly in an erratic manner, so that during any reasonable epoch of observation, the system has visited an enormous number of different microstates. It is therefore appropriate to treat the interaction with the environment as a source of structureless random force acting on the system. This naturally gives rise to a statistical description of the system in terms of the relative frequency of its visiting any of its microstates, and its overall macro-behavior.4

From this account, it is clear that although statistics and probabilities are fundamental to the physical description of the properties of macroscopic systems, this does not imply the abandonment of fundamental determinism at the microscopic level. Once again we can invoke the criterion of predictability in theory. If we had full knowledge of the microstate of the system and its environment at a given time, for example, the positions and velocities of all of the constituent particles, we could, in principle, use the microscopic deterministic laws to predict the microstate of the system and all its physical properties at future times. In other words, what statistical physics does is to distribute the causal powers that operate on a system between its internal forces and the
forces generated by its interaction with the environment. Probability arises when we average out the effect of the environment. However, the fundamental underlying microscopic laws are still fully deterministic.

4. Quantum Indeterminism

The determinist picture of the microscopic physical world was completely overturned by quantum mechanics, a new theory of physics developed at the beginning of the twentieth century. Heisenberg's Uncertainty Principle provides a convenient way of describing what is wrong with classical Laplacian determinism. According to classical mechanics, in order to predict the future position and velocity of a particle, one has to be able to measure its present position and velocity accurately. Heisenberg showed that such a measurement is in principle impossible. The more accurately you try to measure the position of the particle, the less accurately you can measure its speed, and vice versa. In fact, the uncertainty in the position of the particle, multiplied by the uncertainty in its velocity (times the mass of the particle) can never be smaller than a certain universal quantity, known as Planck's constant. Moreover, this limit does not depend on the way in which one tries to measure the position or velocity of the particle or on the type of particle: Heisenberg's Uncertainty Principle is a fundamental, inescapable property of the world.

This observation and other discoveries led Heisenberg, Bohr, Schrödinger, and Dirac in the 1920s to replace classical mechanics by a new theory of microscopic physics—quantum mechanics. This theory abandons the formulation of physical laws in terms of deterministic predictions about the trajectories of particles, replacing them by inherently probabilistic laws. According to quantum mechanics, the future occurrence of a particular microscopic event (such as an electron being in a certain position or in a certain orbit around the atom's nucleus) is unpredictable even if we know with full precision all that can be known about the current state of the system. All that we can deduce from physical laws is the probability of occurrences of each event.5

The fact that according to quantum theory there are no laws that predict conclusively the actual occurrences of physical events is a fundamental revolution in our understanding of the causal underpinning of the physical realm. In fact, for many years, Einstein, one of the early founders of the new physics, was reluctant to accept it as a complete theory of physics because he refused to succumb to the idea that the microscopic underpinning of our universe is indeterminate. The ensuing eighty years of vast experimental and theoretical research have upheld with exquisite accuracy the predictions of quantum mechanics. Furthermore, many of the major technological advances of our era including the transistor, lasers, microelectronics, and more, are based on the results of
this theory. It is therefore not surprising that, nowadays, quantum theory is accepted as the fundamental theory of physics at the microscopic scale.

The emergence of quantum mechanics signaled an end to Laplace’s model of a totally deterministic universe. This does not mean that the world is lawless. Even at microscopic levels, the quantum laws predict the statistical regularity that underlies the frequency of occurrence of events. Most importantly, since the behavior of macroscopic bodies depends on the sum total of an enormous number of microscopic quantum events, most macroscopic properties can be predicted with fantastic accuracy, despite the probabilistic nature of these bodies’ microscopic constituents. In summary, as in the case of thermal noise, quantum fluctuations in the physical states are apparent at the microscopic scale but, in most cases, they are averaged out at the macro level.

III. Neuroscience of Human Choice

1. Neuroscience of Cognition

Ultimately, cognitive neuroscience seeks to link neural processes with mental processes involved in human cognition. In the last decades we have made considerable progress in understanding the neural basis of cognitive functions such as perception, learning, memory, attention, decision making, language, and motor planning.

An examination of the neural correlates of a specific cognitive function can be broken down into several issues. The most basic issue is that of localization: what neural structures in the brain are involved in the function in question, and how does the level of their activation change as the mental process unfolds? Next, there is the issue of representation: how is the activity of nerve cells within a brain area related to the sensory, cognitive, or behavioral variables being processed at a given time? Finally, there is the challenge of discovering the underlying mechanisms: what are the ways through which these representations come about and what is the causal chain that connects them to earlier and to subsequent brain activity patterns?

In animals, studies geared at identifying the brain structures active during specific cognitive processes have used a variety of electrophysiological techniques to measure changes in electrical activity in a brain region. Foremost among these is the technique of recording by microelectrodes. By inserting a thin wire into an animal’s brain, researchers measure changes in the electrical activity of individual cells or small groups of cells. Localization studies in humans use several noninvasive brain-imaging techniques, the dominant one
being functional magnetic resonance imaging (fMRI). This technique effectively determines local enhancement in neuronal activity by measuring the increase in oxygen levels in the blood flow in different brain areas.

Localization data by themselves provide little insight into the nature of the signals encoded in a given structure or the mechanisms that underlie this encoding. To gain further insight, we must know the relation between the activity of neurons in a brain area and the sensory data, the memory, or the behavioral pattern with which this activity is correlated. This defines the neural representation of a brain area. Much of the research in systems neuroscience over the past half century has dealt with uncovering the basic features of neural representations in the different brain areas. Most of this research focuses on the electrical activity at the single neuron level. Using microelectrodes, researchers correlate changes in the electrical activity of individual cells with changes in sensory or behavioral conditions. The results are often expressed in terms of neuronal selectivity. Neurons in sensory areas are selective for specific sensory parameters such as the color of an object, the orientation of a line segment, or the frequency of a sound signal. Neurons in motor areas are active selectively prior to a movement of specific limbs, or even a specific type of movement of that limb. For instance, some neurons in the primary motor cortex of an animal fire strongly before the onset of a planned reaching movement of the arm to a target located straight ahead, but are inhibited from firing before a movement of the same arm to a target located 90 degrees away. Other neurons in the same brain area fire strongly before a movement to the latter target, but will be inhibited by a target located straight ahead of the animal.

As the example above illustrates, different neurons in the same brain area will often have different selectivity or preferences. Researchers can reconstruct the neuronal representation of an area by mapping the different neuronal selectivities within it. This mapping of neural representations is also aided by measuring the effect of stimulating a cluster of cells in a given brain location by injecting electrical current through the microelectrode. Activating clusters of direction-selective neurons in a certain visual area through a microelectrode can induce rhesus monkeys to "report" seeing motion in the direction encoded by the activated neurons. Similarly, microstimulation of localized sites in the motor cortex may trigger a certain arm movement.

But even this level of analysis raises fundamental questions. How are the different neuronal selectivities formed? How are these neuronal patterns of activity deciphered by brain structures downstream? Individual neurons in the cortex typically receive input signals from as many as several thousand other neurons. In a typical experiment, however, the neurophysiologist can characterize the responses of only a few neurons at the tip of an electrode. With such a limited data set, it is difficult to determine exactly how the thou-
sands of synaptic inputs to any given cell are transformed to create the cell’s pattern of output activity. It is thus an enormously difficult challenge to decipher the dynamic mechanisms that underlie neuronal representations in the cortex.

In formulating a theoretical understanding of brain processing, an interesting question is how many neurons participate in a given perceptual or behavioral decision. The difficulty in answering this question stems from the fact that the degree of selectivity of single neurons varies enormously within a given brain area. A small but noticeable minority of cells shows a high degree of selectivity. These cells can be thought of as signaling reliably—by their activity or quiescence—the presence or absence of a particular sensory or motivational variable. Most cells, however, show poor to moderate selectivity, so that the representation of information in any given area most likely resides in the joint pattern of activation and inactivation of a large number of cells, on the order of thousands at least. Furthermore, even the simplest perceptual or behavioral process involves the simultaneous and/or sequential activation of several brain areas, that is, several distinct neuronal representations.

Therefore, although both brain-imaging and single neuron studies often give the impression of a very compartmentalized functional organization of the brain, in fact, even an act as simple as planning a reaching arm movement toward an object involves a complex sequence of activations of large populations of cells in many brain areas: low level visual areas that process features of light patterns on the retina, high level visual areas where objects are detected and recognized, motivational areas where the desirability of an object is stored and assessed, somatosensory areas where the present position of the arm is encoded, and various motor areas where the movement parameters are calculated and executed.

In conclusion, substantial progress has been made in discovering some of the basic anatomical and dynamic components of the neural correlates of cognition. However, neuroscience is still at an early stage of understanding the complex chain of neural events that underlie cognitive functions. Lacking blueprints, discovering function from structure is a vexing task. It is no surprise that modern research aiming at figuring out how the intricate neuronal circuits in the brain generate the appropriate signals is a multidisciplinary enterprise, involving experts from such diverse fields as neurobiology, computer science, physics, and psychology.

It should be stressed, however, that the basic microscopic building blocks of these complex neural circuits are well characterized. Foremost among these are nerve cells: neurons that communicate through a special electro-chemical signaling apparatus—the synapse. Neurons are made of the same “stuff” as other biological cells; their soma contains genetic material embedded in DNA
molecules, their exterior consists of ordinary membrane tissue, and the signals that they transmit take a variety of forms from rapid electrical pulses traversing their membrane (the “spikes”) to the slow secretion and absorption of various chemicals. The important point is that at the microscopic level, the brain does not harbor mysterious or unfamiliar entities. On the contrary, at this level all observed processes are accounted for by the basic laws of physics such as electrochemistry and thermodynamics. Thus, the mystery that the brain poses to us is not due to some indiscernible microscopic features, but rather to the enormous complexity of the organization of the microscopic building blocks. In fact, this challenge is shared by several other domains of contemporary scientific endeavor—understanding the macroscopic properties of systems that consist of many relatively simple microscopic units woven into a complex pattern of organization.

2. Neuroscientific Studies of Decision-Making

Over the past twenty years, neuroscientists interested in vision and other sensory systems have studied the brain areas involved in various levels of sensory information processing. Meanwhile, other researchers have studied the mechanisms behind body movements. Where and how does one look for neural correlates for reasoning, decision-making, volition, and choice? One current working hypothesis is that these should be found in cortical areas that link sensory and motor processing.9

An example of research along this line is the study of the neural correlates of decision-making in monkeys by Platt and Glimcher.10 They placed a monkey in front of a viewing screen and repeatedly presented two visual stimuli—light-emitting diodes on the screen. The animal was trained to direct its gaze to one of the two targets signaled by an appropriate visual cue, in order to receive a juice reward. Then, the experimenters removed the cue signal and let the monkey perform “free choices” between the two targets. In sequential blocks of roughly 100 trials, the researchers varied the size of the juice reward associated with each target. In this case, the monkey selected the two targets with frequencies that were in proportion to their relative reward sizes. While the monkey performed this task, the neural activity in the lateral intraparietal cortex (LIP) was recorded. It was found that given exactly the same visual display and exactly the same gaze shift, neurons in LIP responded differently depending on the animal’s expectation of reward size—an expectation that was evident in the animal’s pattern of decisions. LIP receives inputs from several parts of the brain involved in processing visual information, and it projects directly to the areas that control eye movements. Thus, it may be an important
intermediate stage in the transformation of visual signals into neural commands to move the eyes. As this experiment suggests, this area contains a population of neurons whose activity represents “high level” (e.g., reward expectation) decision-related variables that are neither purely sensory nor purely motor in nature.

3. Brain Structures Underlying Human Choice

Research on human choice cannot be carried out by invasive methods such as microelectrode recordings except under special circumstances. However, the recent advances in brain-imaging mentioned above have invigorated the neuroscientific research on human volition and decision making. The dorsolateral prefrontal cortex (DLPFC), together with brain regions connected to it, has been found to be of particular importance. This area is known to be important for working memory, which plays a crucial role in many cognitive functions. In addition, several studies show that there is an increase in activity in DLPFC when actions (such as word generation and finger movement tasks) are being selected and initiated. Recent brain-imaging studies by Spence and Frith, among others, indicate that the subjective experience of deciding when to act and which action to perform is associated with a characteristic pattern of activity in DLPFC.

It is of interest that patients with symptoms of a “sick will,” that is, inactivity, lack of ambition, autistic behavior, etc., show abnormal activity in the prefrontal parts of the brain. Symptoms of this type are encountered in chronic schizophrenia, depression, and organic dementia. Furthermore, damage to DLPFC in humans leads to stereotyped responses to objects in these individuals’ environment, such as putting on glasses or eating food whenever these are placed in front of them. Such patients might be said to have a major impairment of free will since they have become slaves to their environment.

A direct relationship between electrical signals in the brain and awareness of intention to perform movements was also found in a recent study conducted by Fried and coworkers. Working with epileptic patients, they noticed that electrical stimulation of sites in the frontal cortex caused the patients to report an urge to move a specific body part, or a feeling that they were about to move. Fried’s results are consistent with the view that conscious awareness of intention is conjoined with frontal brain activity.

These and other results indicate that the exercise and experience of choice is correlated with neural mechanisms that are located in the prefrontal cortex and in the brain regions that are linked to it.
4. Lessons from Split-Brain Studies

Important insights into the constraints on human cognitive functions imposed by the anatomy of the brain have been acquired from studies of split-brain patients. In split-brain procedures, done as a last resort in treating certain types of severe epilepsy, the corpus callosum, which connects the two cerebral hemispheres, is surgically cut, and consequently most cerebral communication between the two hemispheres is lost. Split-brain patients superficially function normally in their daily lives, and perform normally in psychometric tests. Nevertheless, careful studies pioneered by the neuropsychologist Roger Sperry have revealed profound changes in these patients’ information processing and decision-making capacities.

In those of us with healthy, intact brains, information presented to our right hemisphere is quickly sent to our left hemisphere, and vice-versa. What happens when the right and left hemispheres of the brain can no longer communicate? Cognitive studies of split-brain patients measure their behavioral responses to stimuli that are presented to only one cerebral hemisphere. Numbers, words, and pictures visually presented to the left hemisphere (i.e., presented only in the right visual field) are verbally repeated or described with no difficulty because the left hemisphere is dominant for language in most humans. Remarkably, such verbal descriptions of sensory input are impossible for the right hemisphere; if an image is shown to the right hemisphere (i.e., presented in the left visual field), the person will be unable to describe it and will usually say that nothing is there. Further experiments revealed varying degrees of transfer of inter-hemispheric information through subcortical pathways. However, this is largely confined to implicit unconscious processes; the explicit conscious system does not have access to this information.

Thus, it can be said that split-brain patients have two largely separate mental systems, each with its own abilities to grasp the environment. Furthermore, amazingly enough, it appears as though some split-brain patients have two systems of will. The best example of this is patient Paul S, whose right hemisphere had considerable language ability prior to the operation. Consequently, researchers were finally able to interview both hemispheres on their views about friendship, love, hate, and aspirations. Interestingly, Paul's right and left hemispheres gave conflicting statements to such questions as what profession he wanted to have, or his attitudes on political events.

The two hemispheres are not only capable of forming their own perceptions and goals but also of generating independent and, at times, conflicting behaviors. One patient found his left hand struggling against his right hand when trying to pull up his pants in the morning. While the right hand tried to pull them up, the left was trying to pull them down.
Interestingly, it was found that the left hemisphere makes major attempts at compensating for its lack of information (in cases where the information was presented to the right hemisphere alone). Michael Gazzaniga showed that when a split-brain patient is subjected to tests in which the left half of his brain does not know the correct answer, it will often make something up based on the information it does have. In one test, each hemisphere was presented with a picture and told to pick the object that relates to that picture. The left hemisphere was shown a chicken claw, while the right viewed a snow scene. The patient pointed to a chicken with his right hand, and to a shovel with his left. After each hemisphere responded, the left hemisphere was asked to explain the choices. When asked what images he saw on the screen, the patient responded, “I saw a claw and I picked the chicken, and you have to clean out the chicken shed with a shovel.”

Trial after trial, this kind of response occurred. While observers of the subject knew exactly why the right hemisphere has made its choice, his left hemisphere could merely fabricate a consistent response. Interestingly, however, the left hemisphere did not offer its explanation as a guess but rather as a statement of fact. Gazzaniga hypothesizes that the left hemisphere of the human brain plays the role of the “interpreter.” While in normal, non-split brains, information from the right hemisphere is delivered to the interpreting mechanism in the left hemisphere across the corpus callosum, in the case of a split brain the left hemisphere will confabulate, inventively integrating the aberrant information into a reasonable story consistent with the left-hemisphere context.

Split-brain research bears several important lessons for the discussion of human reason and will. First, it clearly discredits the notion of a unified holistic entity that is the locus of an individual’s mental capabilities. Here we see that reasoning, will, and the generation of actions can be fractured and compartmentalized within the brain. Cognition and volition are rooted in discrete brain structures that can at times work independently.

Many of our conceptions about the nature of human will are based on our subjective awareness of our own deliberations, reasoning, and conscious choice of action. However, analysis of the reactions of split-brain patients to their own actions serves as a clear warning against taking our subjective intuitions at face value. Our sense of conscious choice may sometimes reflect interpretations that our brain produces about our own actions, rather than factual accounts of how our actions have been actually produced.

5. The Causal Relationship between Willing and Acting

The above experiments provide a general scientific context regarding neural correlates of will and decision making and the constraints imposed on
these faculties by the anatomy of the brain, but do not directly address the question of the causal link between them and behavior generation. Perhaps the most interesting direct attempt to address this question by scientific methods is the famous experiments by Libet and coworkers.18

Previous work has shown that voluntary movements are preceded by a specific wave of electrical activity, which can be picked up by electroencephalogram (EEG) scalp recordings. This precursor brain activity is known as the “readiness potential” (RP). In Libet and coworkers’ experiment, subjects were instructed to perform a simple flick or flexion of the wrist spontaneously at any time they felt the urge to do so. These voluntary acts were performed capriciously, free of any external limitations or restrictions. In these acts, the RPs began on average 550 milliseconds before activation of the involved muscles,19 indicating that the brain had already “decided” to act about half a second before the act itself. The question posed by Libet was: when does the conscious wish or intention to perform the act appear? To address this question, Libet put a rapidly rotating clock in front of the subjects and asked them to mark the position of the clock arm at the time they felt the urge to act. In the traditional view of free will, one would expect conscious will to appear before, or at the onset, of the RP, and thus command the brain to perform the intended act. Instead, Libet found that the subjective awareness of the wish to move appeared some 350ms after the onset of the RP, well after the subject’s decision to move could be predicted by observing brain activity. This experiment has recently been replicated by Haggard and Eimer.20

It would then seem that “the initiation of the freely voluntary act appears to begin in the brain unconsciously, well before the person consciously knows he wants to act!”21 In other words, the conscious feeling of exerting one’s will is an epiphenomenon, “simply a by-product of the brain’s activities but with no causal powers of its own.”22

These findings have generated heated controversy and conflicting interpretations. Libet’s proponents noted that “his experiment brought scientific method to a question that had previously been purely philosophical.”23 Other researchers have voiced strong criticism, raising a number of technical and methodological problems. For instance, reports of conscious experience could be delayed relative to the actual experience itself. Additionally, Libet’s subjects presumably divided their attention between viewing the external clock and their own decision making; hence identifying the time of their conscious will with the reported clock position is problematic.24

The current neuroscientific understanding of human decision-making, choice, and will is still in a very primitive stage, and future progress in noninvasive techniques will surely vastly improve our capability to measure and characterize the underlying neural processes. However, a rough picture has al-
ready emerged: even conscious decisions and intentions to act are consequences of electrical and chemical signals produced within the neural circuitry of the brain. Furthermore, the available physiological data confirm what has been known to psychology for a long time, namely that our conscious awareness of our actions is often not a necessary part of the actual real-time causal chain that leads to their execution. Rather, our awareness is part of the process of evaluation and interpretation of our actions and their consequences. This process forms and molds our representation of the world and of ourselves, a representation that will shape our, perhaps unconscious, future behavior.

It should be realized that the most sophisticated decisions studied in current experiments are still a long way from simple everyday decisions, such as what to eat for dinner. Ultimately, it may be impossible to measure the brain signaling that underlies the complex decision process of the scope that we encounter in everyday life in controlled laboratory settings. However, the general working hypothesis in science is that the constrained processes observed in the lab reveal the underlying mechanisms at work in the more complex settings of the real world. There is no reason to suspect that this is not the case for human cognition and behavior.

IV. Attempts to Reconcile Free Will with Science

1. Metaphysical Free Will

In previous sections, we have surveyed the scientific understanding of the nature of natural laws, and current neuroscientific insights into human volition and choice. In this section and in the following one I will analyze some of the attempts to reconcile this scientific picture with the concept of free will. I will confine myself to one pivotal and highly controversial hypothesis, which is known as the Principle of Alternative Possibilities (PAP), or, put simply, the hypothesis that we have free will in the sense that “we could have done otherwise.” Actually, this notion of free will consists of two hypotheses:

A. Indeterminism of the World

Human choice is a choice between genuinely open alternatives. At the moment of choice, the two options we have in front of us are both possible outcomes. This means that the future is really open. It is undetermined by the present state of the brain and the environment. Given exactly the same state of the world, we could have chosen otherwise, in which case the outcome trajectory of our actions would have been different.
B. Causation of the Will

When a choice is made between two genuine possibilities it is our will that makes these choices. Our conscious free will has a genuine causal power to intervene in the physical world and determine which of the available possibilities will be realized.

For the sake of brevity, I will call this notion of free will “metaphysical free will.” It is often denoted as an incompatibilist conception of free will. In contrast, various compatibilistic forms of “practical free will” deny that we choose between genuinely open alternatives. These notions of free will in general do not run counter to scientific ideas or findings. It is the strong sense of free will, the metaphysical free will, which stands in apparent conflict with scientific world-view, as will be explained in the following sections.

2. Quantum Indeterminism and Free Will

Since the study of quantum mechanics began, there have been recurrent attempts to explain human freedom by appealing to quantum indeterminism. Sir Arthur Eddington, a prominent British astronomer and philosopher of science at the turn of the twentieth century, was among the first to speculate that Heisenberg’s Uncertainty Principle might explain the mind-body interaction. He suggested that quantum fluctuations in the position of synaptic vesicles in the presynaptic nerve fiber, which contain the synaptic neurotransmitter molecules, might influence the discharge of these molecules. This could influence the firing of the post-synaptic cell and trigger “an unstable cascade,” leading to a global change in brain state.26

A more recent advocate of the link between freedom and quantum mechanics is the Nobel laureate neurophysiologist John Eccles, who proposed a detailed theory of mind/brain interaction based on quantum mechanics. Eccles’s starting point is the well-known fact that synaptic release of neurotransmitters does not follow deterministic law but seems to be a stochastic event. The conventional wisdom is that the fluctuations in synaptic events are due to thermal noise. In contrast, Eccles proposed that the fluctuations in synaptic release are quantum in origin, reflecting the indeterministic transitions of the presynaptic membrane between two quantum states. Second, according to this theory, the quantum probability of release of a vesicle can be momentarily changed by mental events without infringing on the law of conservation of energy. But how does the spiritual mind influence the physical processes in the brain? Eccles posited the existence of nonphysical units of mental intention called “psychons,” which act through a quantum probability field to affect the probability of synaptic transmission during a brief period.27
The idea that quantum indeterminism enables the freedom that underlies our will is also shared by several contemporary philosophers. For instance, Hilary Putnam argues that a necessary condition for freedom of the will is indeterminism of physical laws, and that this indeterminism is provided by quantum physics. John Searle argues in a recent article that “all indeterminism in nature is quantum indeterminism; consciousness is a feature of nature that manifests indeterminism”; hence the conclusion that “consciousness manifests quantum indeterminism.” Few physicists today defend the view that quantum mechanics is intimately related to the phenomenon of consciousness or free will. Notable among them is the mathematical physicist Roger Penrose. His theory of consciousness is based on some esoteric and highly dubious speculations on both quantum theory and brain processes.

3. Free Will as an Emergent Property of the Brain

Eccles’s theory is an example of a dualist theory, as it relies on the presumed presence of a nonphysical entity (substance or force) that directly influences microscopic processes within the brain. Dualist theories are rare in current scientific and philosophical thinking. Many contemporary attempts to formulate a coherent theory of free will rely on the concept of emergence. In contrast to the dualist explanations for the mind-body interaction, emergence theories of mind-body interactions are committed to a physical conception of causation.

Emergentism was originally articulated at the beginning of the twentieth century by C. Lloyd Morgan, Samuel Alexander, C. D. Broad, and others, and has regained considerable attention in the past few years, as a reaction to straightforward reductionist trends in science generally and in the study of the mind in particular. Neuropsychologist and Nobel laureate Roger Sperry, for instance, wrote in favor of emergentism as an alternative to reductionism in the 1980s. “To attempt to explain an entity in terms of its parts and then the parts in terms of their parts and so on, results in an infinite regress in which one is left at the end trying to explain everything in terms of next-to-nothing. At each step of the way critical pattern components of causality are lost and explanation becomes less and less complete at each lower level.”

According to Sperry, when a new entity appears, the properties of the entity as a whole overpower the causal forces of its component entities at all the successively lower levels. The new properties of the system (instead of its own original properties) are the ones that determine its eventual fate. Although the causal forces at the microscopic level continue to operate, they are overwhelmed by the new causal properties that emerge in the whole. Thus, writes Sperry, “instead of a universe completely controlled by quantum mechanics and the basic forces of physics, science presents, by this interpretation, a universe controlled
by a rich profusion of qualitatively diverse emergent powers that become increasingly complex and competent.” In the brain, controls at the physico-chemical and physiological levels are superceded by new forms of causal control that emerge at the level of conscious mental processing, where causal properties include the contents of subjective experience. Causal control is thus shifted in brain dynamics from levels of pure material determinacy to levels of mental or subjective determinacy. The flow of nerve impulse traffic and related physiological events are no longer regulated solely by events in kind, but are moved by the higher mental controls. These conscious subjective properties are interpreted by Sperry to have causal potency in regulating the course of brain events; that is, the mental forces or properties exert a regulative control influence in brain physiology. The subjective forces are thus realities that supercede microscopic laws. “Just as the programming variables of a TV monitor have to be included in order to account for the electron flow pattern of the system, so also in the brain the subjective, mental variables of cerebral function have to be included to give a full account of the flow patterns of neural excitation.”

The idea that human choice is an emergent phenomenon and can therefore exert causal influence on neural events has been adopted in one way or another by several thinkers. John Searle, who was quoted earlier, attempts to combine quantum indeterminism in the microscopic domain of neural events with the emergent causal forces of consciousness that fill the “quantum gap” and act to determine the course of events in the brain. Somewhat analogously to the concept of emergence, a current trend in philosophy of mind utilizes the concept of “agent causation.” According to the proponents of this view, “agents,” such as human beings, have fundamental causal powers well above the familiar microscopic physical causal forces.

The general idea that emergent phenomena are to some degree detached from the underlying microscopic forces and in turn can have causal influence on the microscopic events within the system has recently attracted considerable interest among philosophers attempting to redefine part-whole and inter-level relationships in complex systems. Emergence, in their view, is a conceptual framework for nonreductive materialism. They are committed to a scientific-materialistic world-view but argue against the idea that every phenomenon is reducible or predictable even in principle, by the forces acting on its microscopic constituents.

V. Human Choice and the Microscopic Closure of the Physical World

Having presented the main attempts to integrate the concept of free will with a scientific outlook, I present a critical appraisal of them below.
I begin by addressing the question whether, given our current understanding of neural processes, it is plausible that the brain does not follow strict deterministic physical laws.

1. Is the Brain an Indeterministic System?

One of the most prominent characteristics of the responses of single neurons to sensory stimuli is the apparent random nature of these responses. Specifically, the number and timing of electrical pulses ("spikes") that are emitted by individual cells in the cortex following a sensory stimulation vary considerably across repeated presentations of the same stimulus.\textsuperscript{35} Additionally, as mentioned above, the signal transmission through synapses, which is the main form of communication within the brain, also exhibits considerable randomness. Often, synapses may completely fail to transmit the electrical impulse of the presynaptic cell to the postsynaptic cell.\textsuperscript{36} These are just two examples of well-documented apparent randomness in the way neurons in the brain respond to events in the external world. It is therefore not surprising that most current theories describe the mapping between sensory and behavioral events and neuronal electrical activity patterns in probabilistic terms.\textsuperscript{37} From this perspective, our current understanding of the brain supports the view that brain processes are appropriately described by indeterministic, statistical laws.

The origins of the ubiquitous stochasticity of neural signals are only partially understood and are still the subject of experimental and theoretical investigation. It is, however, clear that the main sources of neural noise are forces that can be characterized as thermal or chaotic rather than quantum in nature. They range from the thermal fluctuations in the opening and closing of ionic pores in the cell membrane\textsuperscript{38} to the fluctuations in the signals that impinge on the sensory organs in ordinary ambient conditions. The complexity of the dynamics of neuronal circuits in cortex may give rise to chaotic patterns of activity,\textsuperscript{39} thereby providing additional source of irregularity in neuronal signaling.

As explained in Section II, thermal noise and chaos are all deterministic in the sense that the future trajectory of brain states is completely determined by the present state of the brain and the environment. We may observe that upon repetition of the same stimulus, the response of a nerve cell varies. This is because neither the initial condition of the neuronal networks in the brain nor the environment surrounding the brain remain the same at each stimulus presentation. These differences are responsible for the observed variations in the neuronal responses. If there is nothing more than thermal noise and chaos at play, were we able to recreate exactly the same conditions for the brain and surrounding, the outcome of each trial would not vary. Thus, neither thermal
neural noise nor chaotic brain dynamics give rise to an indeterminism of the nature required by the premise of "metaphysical freedom."

Does quantum indeterminism play a significant role in brain function? Our present understanding of synaptic biophysics does not support the specifics of Eccles's theory regarding the quantum nature of synaptic transmission. More generally, the electrical currents in the brain are made of quantum elements and processes, like the rest of the universe. Therefore quantum fluctuations on the atomic or electronic level will leave some minute imprint on the noisy environment of the neuron. Since, in most circumstances, quantum noise is negligible in magnitude compared to thermal noise, quantum effects probably do not play a noticeable role in the probabilistic laws underlying brain function. On the other hand, it cannot be ruled out that quantum effects do meaningfully influence the particular neuronal firing times of single neurons in individual events, as discussed below.

For quantum noise to affect brain processes, the relevant quantum fluctuations must be noticeable in single nerve cells, on length scales of at least several micrometers and time scales of at least several milliseconds. These scales are many orders of magnitude larger than those of individual quantum events. Furthermore, as discussed in Section III, ordinary behavioral decisions are expected to involve not just a few neurons, but rather the summed activity across large neuronal ensembles, putting additional constraints on the plausibility of the thesis that quantum indeterminacy plays an important role in brain function. The gap between the microscopic quantum scales and macroscopic neuronal activity may be surmountable if the neural system is extremely sensitive to minute changes in the forces acting on it. In fact, such sensitivity is expected if neuronal circuits in the brain are highly nonlinear chaotic systems. Chaos within the brain may amplify enormously the small quantum fluctuations at the atomic scale to a degree that will affect the timing of spikes of neurons in the brain circuits. This might be relevant in particular cases where choice is being made between alternatives with nearly equal "weights." In such borderline cases, small fluctuations in firing patterns of a group of cells may tilt the choice in favor of one of the alternatives.

In sum, given the present state of our understanding of brain processes and given the standard interpretation of quantum mechanics, we cannot rule out the possibility that the brain is truly an indeterministic system; that because of quantum indeterminism, there are certain circumstances where choices produced by brain processes are not fully determined by the antecedent brain process and the forces acting on it. If this is so, then the first prerequisite of "metaphysical free will" (Hypothesis A, Section IV) may be consistent with the scientific understanding of the brain. Before I discuss the second prerequisite
(Hypothesis B), I present below a critical analysis of the concept of *emergence* and its application to the brain.

2. Is Human Choice an Emergent Property of the Brain?

The idea that large complex systems have *emergent* macroscopic properties is a cornerstone of our contemporary understanding of physical phenomena. The simplest examples of emergent properties are the macroscopic states of ordinary matter, such as gaseous, liquid, and solid states. The very distinction between gas and liquid, for example, applies only to large systems of atoms or molecules; it is meaningless to speak of a single molecule or even a small group of molecules as being in either a gaseous or liquid state. We thus can say that gaseous and liquid states, viscosity, solidity, and the like, are all emergent features of macroscopic systems.

The concept of emergence is essential not only to physical laws of the macroscopic domain. Current theories of microscopic physics are also based on the notion of emergence. For instance, the present Standard Model of elementary particles is an incomplete theory of matter on the smallest possible spatial scale. For one thing, it does not include gravitational forces between elementary particles, which would play an important role below an extremely short distance known as the Planck length. Furthermore, the mathematical structure of the Standard Model is unintelligible unless this theory is viewed as an effective theory describing physics at the intermediate scales of the subatomic phenomena. We therefore must treat the very notions of electrons and quarks not as describing fundamental units of nature, but as "emergent" features of a more fundamental theory, as yet undiscovered, which describes the appropriate physics at the fantastically small Planck scale. Theories in physics at all levels are cast in terms of entities that are emergent features at the appropriate level.

The relationship between emergent properties and the underlying microscopic laws is an intricate one and the underlying principles have surfaced only in the last thirty years. On the one hand, emergent properties are largely insensitive to the detailed configuration of microscopic constituents. This is, of course, the reason why we can rely on macroscopic laws of matter. The great triumph of theoretical physics, the above-mentioned Standard Model, is exquisitely accurate in describing the interactions between quarks, electrons, and photons, despite the fact that we do not yet have a good idea what physics at the "Planck scale" looks like! On the other hand, emergent properties do depend on some aspects of the microscopic states. For example, the freezing of a liquid depends on the mass of the constituent molecules, and the nature of the forces between them. Thus, emergent properties are independent of many
“irrelevant” parameters of the microscopic systems, but do depend on a small set of “relevant” microscopic parameters. Although this principle is nowadays well understood, the actual determination of which parameters are relevant for any given emergent phenomenon may be extremely difficult.

The electrical signaling of nerve cells operates on length and time scales that are large compared to the underlying molecular processes. Furthermore, typical cognitive functions involve complex interactions between many nerve cells distributed across several cortical and subcortical areas. It is natural to consider them as emergent properties of the microscopic brain processes. The view that cognition and other mental phenomena are emergent macroscopic features of brain processes is bolstered by the substantial level of neuronal noise, which, as discussed above, will anyway “average out” the details of the microscopic state of the brain in any given instance. Which aspects of the neuronal firing patterns are relevant for any emergent cognitive faculty is still largely unknown, and remains a fundamental problem in brain theory.

3. Free Will and the Principle of Microscopic Closure

We are now ready to address the question of whether our current scientific understanding is consistent with the notion of metaphysical free will. As described in Section III, we have abundant and decisive scientific evidence that mental faculties such as volition, intention, and decision are rooted in the underlying patterns of neuronal electrical activity in the brain. Yet, can we identify free will as a causal force that works by influencing the neuronal activity in the brain? Let us begin with the “quantum theory” of free will. Above, I have argued that the hypothesis that the brain is affected by quantum fluctuations in a manner that has some consequences for choice and decisions is not inconsistent with our current understanding of the biophysics and dynamics of the underlying neuronal circuitry. However, as described in Section IV, metaphysical free will requires not only an element of indeterminism but also the existence of an entity, our will, that influences the course of events and fills in the indeterministic gap between events. It is this premise, that we have a will with its own causal power to interact with the physical processes in the brain, that stands in apparent contradiction to our scientific understanding.

According to the theory of Eccles and others, our will is efficacious by virtue of some nonphysical entity or force that intervenes at the molecular levels of, say, synaptic transmission and modifies the quantum probabilities associated with this process. If mental forces result in event probabilities in the brain that are different from those predicted by the known physical laws, these violations of the laws should be readily detected and verified by ordinary scientific observation. As discussed in Section III, there is overwhelming scientific evi-
dence that the microscopic processes in the brain are just those predicted by familiar laws of physics and chemistry. Thus, such a hypothesis stands against all current scientific evidence. Furthermore, if violations of the known laws of nature do occur in the brain, then following their detection by scientific methods, we would necessarily change our theory of nature and incorporate these novel forces as part of the scientific description of the natural causal network. We will need to characterize the spatial and temporal properties of the new forces, their causal chains, and their relation to the causal chains of other physical forces, etc. In sum, these new entities, if they exist, will have to be integrated into the framework of science as part of the physical forces acting in the brain. This raises a serious problem with regard to the very notion of a nonphysical, mental entity that intervenes regularly in microscopic brain processes in violation of the regularities imposed by physical forces.

Let us now turn to the idea of free will as an emergent property of brain processes. As we have discussed, the notion that will and choice are emergent features of the brain is a scientifically sound and appealing idea. Furthermore, it would seem that emergence presents a straightforward resolution of the difficulty posed by the interaction between the mental and the physical. This is because according to emergence theory, the mental does not represent some mysterious new microphysical force. In fact, virtually every complex system exhibits emergent properties. Construed as such, we might argue that our will is as physical as the solidity of a solid or the fragility of a glass.

The major problem with the emergence explanation of free will is that it is based on the notion of “downward causation.” To explain the causal powers of free will, emergentists assume that because the will is part of the emergent macroscopic realm of the brain, it exerts a “regulative control” on its state. However, since action ultimately requires the appropriate neural electrical signals, the controlling power of the will must come about by its capability to influence these microscopic processes. Indeed, the emergentists hold that even in general complex systems other than the brain the emergent properties influence their microscopic state, working against, or at least beyond, the microscopic causal powers.

While the notion of emergence is familiar to contemporary science, the idea that emergent properties have independent causal powers is completely foreign to it. Emergent features of a macroscopic system are determined by nothing other than the microscopic constituents of the system, their spatial arrangement and the microscopic forces acting between them. What is special about emergent features is, first, that they represent outcomes of the microscopic physical processes that appear only on the macroscopic scale, and second, that they bear very nonadditive and complex dependencies on the specific details of the microscopic states and force parameters. Nevertheless, they
do not assume causal relations that are not rooted in the underlying microscopic laws.

Emergence theories of metaphysical free will often cite analogies from ordinary physical systems to illustrate the concept of downward causation. Earlier, I cited Sperry’s example of the programming variables of a TV monitor as being essential in accounting for the electron flow pattern of the system. Another familiar example used by Sperry is a wheel rolling down a hill; the molecules of the wheel rotate by virtue of the wheel’s macroscopic motion. However, none of these analogies actually support the concept of downward causation. The rotation of the wheel could also be seen as either a consequence of the motion of the constituent molecules or (depending upon the precise definition of terms) as equivalent to the organized motion of the molecules. Given our understanding of the laws of motion, it does not make sense to consider the motion of the wheel as causing the motion of its molecules. Likewise, the design of a TV monitor does not exert a force on the electrons inside it beyond the usual microscopic forces, but rather utilizes these forces. The device is configured precisely so that the electromagnetic forces between electrons in the TV tube will accelerate them, according to the known laws of motion, to the desired spots on the fluorescent screen. By introducing an element of intelligent design, the example of the TV monitor also confuses the teleological reasoning commonly used in explaining human action with physical mechanisms acting as causes of events.43

John Searle acknowledges the flaw in the above analogies. In defining the nature of consciousness he argues:44 “The trajectory of each molecule is affected by the behavior of the entire solid wheel. But of course there is nothing there but molecules. The wheel consists entirely of molecules. So when we say the solidity functions causally in the behavior of the wheel and in the behavior of the individual molecules that compose the wheel, we are not saying that the solidity is something in addition to the molecules; rather, it is just the condition that the molecules are in.” Likewise, “a system feature [e.g., consciousness] can affect micro-level elements in a system composed entirely of the micro-level elements [e.g., neurons], in which all causal powers are reducible to the causal powers of the micro-level elements” (my italics). This account, however, does not leave room for real causal powers of the will, if indeed viewed as just a macro-feature of the brain.

As quoted in Section IV, Searle goes on to argue that free will is different from the motion of the wheel because, in contrast to the wheel, consciousness is affected by quantum indeterminism. It is difficult to apprehend how Searle’s argument resolves the inconsistency of the macro/micro relationship. After all, if the macrostate of the brain is nothing but the collective state of its micro-elements, how can this macro state influence neuronal electrochemical processes beyond what is determined by the microscopic laws? Sim-
ilar objections apply to attempts by other philosophers of the Libertarian camp to defend the concept of metaphysical free will by positing some sort of "agent causation." 45

Finally, we address the hypothesis that an act of free will should not be considered in terms of physical laws acting either on the micro or on the macro levels of the brain. Rather, it should be considered as a nonphysical (mental or spiritual) influence on brain processes that does not violate physical laws but supplements them. For example, perhaps our will intervenes at the quantum level and determines the *actual realization* of quantum events in the brain but *does not change the quantum probabilities* of the different events. This way, the argument might go, free will acts in complete consistency with the regularities predicted by ordinary physical laws. It just fills in the indeterminacy gap left by them. According to this conception, our free will is severely restricted in its scope of influence as it is constrained by the statistical regularities imposed by natural laws. Furthermore, if the mental influences on behavior are fully consistent with physical laws, it follows that all observable features of human cognition and action can, in principle, be explained on the basis of physical processes within the brain. In such a case, from a scientific perspective, invoking putative nonphysical entities, such as free will, in explaining human behavior is redundant.

What emerges from the above objections is a fundamental scientific view of the causal underpinning of all physical phenomena, which stands in contradiction to the incompatibilist conception of "metaphysical free will." I term this view:

The microscopic closure of the physical world: *The only factors that cause or influence physical events are those that result from (or are reducible to) the physical properties of the constituent microscopic elements and the microscopic forces acting between them.*

Note that this hypothesis does not rule out the existence of undetermined or random events in nature. It merely states that whatever is not determined by the physical forces operating at the microscopic level cannot be determined by something else. On the face of it, the hypothesis of the microscopic closure of all physical phenomena seems to be presumptuous. What is the basis of our claim that physical events cannot be determined by other influences, particularly if these influences work within the *statistical regularities* dictated by the physical laws? The basis for this claim is rooted in the foundation of the scientific conception of the physical world. Science provides us with a framework of describing causal chains between physical events mediated by physical laws. According to this framework, any change in the motion or state of particles (such as that which must occur in the brain prior to any action) is by
definition a physical force. As such, its effects are subject to the scientific methods of empirical verification and to tests of their consistency with other known natural laws and principles. As explained above, if there are observable deviations from known natural laws, these deviations will be indicative of the presence of hitherto unknown physical forces and will be integrated into an updated physical theory of the brain. If, on the other hand, the nonphysical entities do not cause any observable deviations from known physical laws, then admittedly the physical laws are fully adequate in explaining all observables including human action. Thus, the scientific conception does not provide us with any coherent framework for describing nonphysical influences on physical systems.

The above “microscopic closure” hypothesis is related to the more familiar hypothesis of the “causal closure of the physical domain.” The version presented here emphasizes the exclusion of emergent scenarios of downward causality. As mentioned above, no example of emergence understood scientifically exhibits downward causation. They are all predictable from, or reducible to, microscopic causal laws. In fact, it is a fundamental character of the scientific conception that complex phenomena ought to be explained in terms of laws that are derivable from the underlying microscopic level. Whether the concept of downward causation is coherent at all is questionable. Even if philosophers were able to explicate it, it is unclear how a scientific conception of reality could incorporate it any more than it could incorporate the notion of nonphysical forces acting at the microscopic level.

VI. Conclusion

I was asked to address the issue: “What measure of non-biologically determined behavior can coincide with current neuroscientific conceptions of the role of physiology in determining behavior?” What is then my answer to this question? Well, it depends crucially on the meaning of the term “non-biologically determined behavior,” and, to some extent, also on what we take to be our “current neuroscientific conceptions” of neuronal mechanisms underlying behavior. As I have outlined in this paper, given the current state of understanding of the brain, neuronal activation and the spread of this activity is influenced by a substantial degree of noise. Hence, we expect that there will be many cases where the chains of events linking stimuli to human cognition and behavior are appropriately described by statistical laws rather than deterministic ones. In particular, such laws would allow, and even predict, that upon recurrence of similar external conditions, our actions might take a different course, due to the small variations in the internal neuronal
states as well as in the external forces in constant interactions with the nervous system.

On the other hand, I have argued that both external and internal sources of noise are by and large "classical" in nature. This means that although their realization is unpredictable in practice, they are fully determined by the antecedent condition of the system and the forces acting on it. In this sense, these processes are still biologically determined. More specifically, in the context of the problem of freedom of choice or free will, given exactly the same internal and external conditions we would not have been able to act differently from the way we did. Still, I have argued that there might be conditions in which the effect of quantum indeterminacy on neural processes is amplified so that it has a noticeable effect on behavior. This would be the case if the relevant neuronal circuits exhibit strongly chaotic dynamics, which would make them extremely sensitive to small quantum fluctuations. In these cases, the particular behavioral outcome of the brain process would be truly undetermined by previous conditions and processes. Assessing the feasibility of this scenario must await more theoretical and experimental research. In any case, we expect this to be the exception, obtaining, for instance, only in cases where the nervous system is near an unstable equilibrium or extremely close to a threshold, rather than the rule.

Finally, I have addressed the incompatibilist conception of metaphysical free will. I have discussed two main lines of defense of this notion of free will. One relies on hypothetical mental influences directing quantum microscopic fluctuations. The other assumes that the emergent features at the macrolevel of complex systems have causal powers to influence future macro- and micro-events (downward causation). I have argued that, from a scientific standpoint, even if certain physical events are not fully determined by the underlying physical laws, there remains no good explanation as to how they can be influenced by other factors, operating either at the micro or macro levels. Thus, even if certain behaviors are not fully determined by prior biological processes, we have no explanation for how they might be affected by an extraphysical choice or will, namely by a power that would not effectively fall under the scientific character of a physical force.

"Any philosopher"—writes philosopher Daniel Dennett—"ought to feel at least a little embarrassed that with so much work so little progress has been made." Searle agrees. "The persistence of the traditional free will problem in philosophy," he writes, "seems to me something of a scandal." I am not qualified to judge whether these assertions are just. However, it does seem to me something of a scandal that scientists and scientifically literate philosophers who defend metaphysical freedom of the will pay so little attention to the classical problem of interaction between the will and our brain processes. Since
the dismal attempts of Descartes to resolve this difficulty there has not been a single scientifically sound hypothesis that explains how our behavior can be determined by an entity not governed by physical laws.

Our objection to incompatibilist metaphysical free will does not imply that human will and choice are unimportant epiphenomena. It simply implies that they are emergent features of the causal chain of physical events that occur within our brain. Delineating the role of reasoning and choice in human behavior remains one of the most important challenges of science as well as philosophy, moral and religious thought.

"Man," writes Einstein, "defends himself from being regarded as an impotent object in the course of the Universe. But should the lawfulness of events, such as unveils itself more or less clearly in organic nature, cease to function in front of the activities in our brain? . . . Determinism does not stop before the majesty of our human will."

"If physical determinism is true," writes Eccles, "then that is the end of all discussion or argument; everything is finished. There is no philosophy. All human beings are caught up in this inexorable web of circumstances and cannot break out of it. Everything that we think we are doing is an illusion and that is that. Will anybody live up to this situation?"

Here are two divergent reactions of distinguished scientists to the prospect that human will and choice, reason and action, all succumb to the deterministic laws of nature. It is time to adopt the modesty preached by Einstein and to give up Eccles's fatalistic worries. We, human beings, are the magnificent product of God's nature. This is not "the end of all discussions." It is merely the end to the wrong discussions, mistakenly putting man in the place of God. This should be the beginning of the right discussions: how to sustain humanity and human values within the natural order decreed by Him.

Notes

I am grateful to Dr. Ruty Lev Bar-Or for her help in writing this article, to Dr. Yuval Dolev for useful discussions on the subject matter, and to Olivia White for helpful and illuminating comments on the manuscript. I also thank Dr. Yitzhak Berger for useful suggestions. I acknowledge the support of the Center for Systems Neuroscience, Harvard University, where the final version of the paper was written.


4. I have presented an interpretation of statistical physics in terms of interaction with the environment. This interpretation is the most natural in the context of biological systems. Under certain conditions, statistical physics can also be justified for *isolated* physical systems. In either case, classical statistical physics is fully consistent with classical micro-determinism. See for example, Sklar, Philosophy of Physics, ch. 3.

5. Actually, in the "standard" interpretation of quantum mechanics, also known as the "Copenhagen interpretation," the quantum laws do not even predict the probabilities of the occurrences of events but only the probabilities of the outcomes of their measurement. This raises difficult fundamental problems about the role of the observers in physical laws and the definition of measurement. In recent years, there has been a resurgence of attempts to rid quantum theory of the notion of measurement and even to reconcile quantum theory with classical determinism. See Sklar, Philosophy of Physics, ch. 4; Sheldon Goldstein, "Quantum Theory without Observers," Physics Today, March 1998: 42–46, and April 1998: 38–42.


19. The time of onset of movement is measured by the onset of electrical activation of the muscles (EMG). In some experiments, RPs precede the onset of movement by as much as one to two seconds. See refs. in notes 21–22 for detailed discussions of this issue.


21. Benjamin Libet, "Do We Have Free Will," in The Volitional Brain, 49.


25. PAP figures prominently in debates on moral responsibility. In this article I am referring to the implication of this principle to the very notion of free will. The implications on the concept of moral responsibility are beyond the scope of this article. For discussions of PAP, see David Shatz, "Irresistible Goodness and Alternative Possibilities," in Freedom and Moral Responsibility: General and Jewish Perspectives (College Park: University Press of Maryland, 1997), ed. C. H. Manekin et al., and other articles in this volume.


33. See Searle, note 30.
41. As is well known, the gravitational force between two masses falls off at the inverse of the squared distance between them. At large distances, this force is strong only among objects such as the earth and the sun due to their huge masses. With microscopic particles this force is negligible compared with electromagnetic and nuclear forces, except at extremely short distances—at the scale of Planck length, which is many orders of magnitude smaller than the distances probed by present day most powerful accelerators. See refs. in note 41.
43. Unless intelligent reasoning is assumed to have real causal powers independent of the neural processes in the designer’s brain. This, however, is the very issue in question.
44. See note 30.
45. See the criticism of such “extra factors” by Robert Kane, “Reflections on Free Will, Determinism and Indeterminism,” www.ucl.ac.uk/~uctytho/dfwVariousKane.html. Kane, himself a libertarian, argues that free will is exercised only on rare occasions when we make difficult choices between several equally reasonable alternatives; these choices affect our beliefs and motivations and thereby our future deterministic choices as well (hence they are called Self-Forming Alternatives [SFAs]). See Robert Kane, The Significance of Free Will (Oxford, 1996). Kane seems to hold that choices in SFAs are undetermined because of an unstable equilibrium in the underlying neuronal dynamics.
However, the actual choices are the outcome of “decisions” made by internal neuronal noises and not by “external” causal powers. Nevertheless, since the equilibrium itself and the ensuing process of deliberation involve the agents’ efforts, beliefs, goals, etc., they are justifiably called the choice of the agent’s will, as long as the outcomes of the process are, in fact, compatible with his will. Thus, in the present context it seems that Kane is satisfied with a definition of free will that entails indeterminism (Hypothesis A, Section IV) but not the causal powers of the will in the usual sense (Hypothesis B).

46. Observable deviations may take the form of systematic laboratory recordings of microscopic processes, for example, the detection of novel electrical properties of nerve cells, or systematic psychological observations of behavioral patterns with regularities that are inconsistent with the physically based theory.

47. See Jaegwon Kim, note 32.


49. This is somewhat akin to Kane’s SFA, see note 46, and David Shatz, ref. in note 26.

